

PREDIKSI KELULUSAN MAHASISWA DENGAN METODE *SUPPORT VECTOR MACHINE* (SVM) STUDI KASUS STMIC CIKARANG

Tumini

Program Studi Sistem Informasi STMIC Cikarang

E-mail : kemuningijo@gmail.com

ABSTRAKSI

Tujuan pendidikan nasional adalah untuk menghasilkan lulusan yang menguasai cabang Ilmu Pengetahuan dan/atau Teknologi untuk memenuhi kepentingan nasional dan peningkatan daya saing bangsa. Tolok ukur keberhasilan sebuah perguruan tinggi adalah tingkat kelulusan dari mahasiswa yang tinggi. Masalah yang banyak dihadapi oleh banyak perguruan tinggi saat ini adalah rendahnya tingkat kelulusan mahasiswa. Salah satunya adalah Sekolah Tinggi Manajemen Informatika dan Komputer (STMIC) Cikarang. Prediksi kelulusan mahasiswa menjadi hal yang penting dilakukan oleh perguruan tinggi berdasarkan data lampau yang dimiliki oleh perguruan tinggi. Penelitian prediksi kelulusan mahasiswa di STMIC Cikarang ini menggunakan metode CRISP-DM dalam pembuatan prototype Aplikasi Data Mining. Langkah-langkah CRISP-DM adalah Fase Pemahaman Bisnis, Fase Pemahaman Data, Fase Pengolahan Data, Fase Pemodelan Fase Evaluasi. Dalam pemodelan data mining menggunakan metode Support Vector Machine (SVM). Support Vector Machine (SVM) mencari hyperplane terbaik yang memisahkan dua buah class pada input space. Proses mining terdiri dari proses training dan proses testing menggunakan data dari sistem informasi akademik. Untuk evaluasi dan validasi hasil pemodelan data mining menggunakan metode Cross Validation dan Kurva Receiver Operationing Characteristic (ROC). Hasil dari penelitian ini adalah model datamining yang dibuat dengan rapidminer dan aplikasi datamining yang dibuat dengan bahasa pemrograman C#.

Kata Kunci : CRISP-DM, Data mining, prediksi, SVM, UML

1. PENDAHULUAN

Pendidikan Tinggi merupakan kelanjutan pendidikan menengah yang diselenggarakan untuk menyiapkan peserta didik menjadi anggota masyarakat yang memiliki kemampuan akademik dan profesional yang dapat menerapkan, mengembangkan dan menciptakan ilmu pengetahuan, teknologi dan atau kesenian. Program yang ada dalam pendidikan tinggi ini tidak hanya sarjana (S-1) melainkan diploma, pendidikan profesi, magister (S-2), bahkan doktor (S-3). Sedangkan satuan pendidikan yang menyelenggarakan pendidikan tinggi ini dikenal dengan nama Perguruan Tinggi (PT), baik itu Perguruan Tinggi Negeri (PTN) maupun Perguruan Tinggi Swasta (PTS).

Salah satu tujuan dari pendidikan tinggi yang termaktub dalam Undang-Undang UU No. 12 Tahun 2012 tentang Pendidikan Tinggi yaitu pada pasal 5 adalah Dihasilkannya lulusan yang menguasai cabang Ilmu Pengetahuan dan/atau Teknologi untuk memenuhi kepentingan nasional dan peningkatan daya saing bangsa.

Lulusan juga menjadi poin penilaian dalam proses akreditasi Program Studi. Standar 3 Borang 3A mengharuskan program studi untuk menampilkan data mahasiswa baru dan lulusan pada angkatan tersebut.

Lembaga pendidikan menghadapi masalah siswa yang terkait dengan tingkat kelulusan, perguruan tinggi dengan mahasiswa lebih tinggi tingkat retensinya dan cenderung memiliki tingkat kelulusan yang lebih tinggi dalam waktu empat tahun. Rata – rata tingkat retensi nasional mendekati 55%.

Hal yang sama juga dihadapi oleh Sekolah Tinggi Manajemen Informatika (STMIC) Cikarang. Dalam kurun waktu 2006 sampai dengan 2010, tingkat kelulusan mahasiswa dibandingkan dengan pendaftaran mahasiswa baru menunjukkan nilai yang belum sesuai dengan harapan. Tabel I-1 menggambarkan kondisi tersebut. Menjadi sebuah kebutuhan untuk bisa mendeteksi mahasiswa yang berpeluang untuk tidak lulus dengan berbagai alasan.

2. LANDASAN TEORI

2.1 Definisi Data Mining

Berikut ini adalah beberapa definisi data mining dari berbagai sumber :

1. *Data mining*, sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. (Santoso, 2007)
2. *Data mining* adalah proses menemukan korelasi baru, pola dan tren dengan memilah-milah sejumlah besar data yang disimpan dalam repository menggunakan teknologi pengenalan pola serta statistik dan teknik matematika. (Larose, 2005)
3. *Data mining* diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data

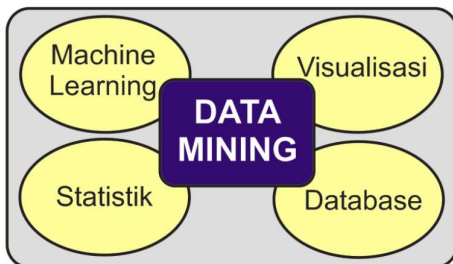
besar yang membantu dalam pengambilan keputusan. (Prasetyo, 2012)

2.2 Posisi Data Mining dalam berbagai Disiplin Ilmu

Data mining sering dikaitkan dengan pengenalan pola (*pattern recognition*), statistik dan *Machine learning*. Pengenalan pola adalah suatu disiplin ilmu yang mempelajari cara-cara mengklasifikasikan obyek ke beberapa kelas atau kategori dan mengenali kecenderungan data (Santoso, 2007)

Machine learning adalah suatu area dalam *artificial intelligence* atau kecerdasan buatan yang berhubungan dengan pengembangan teknik-teknik yang bisa diprogramkan dan belajar dari data masa lalu. (Santoso, 2007)

Pengenalan pola, *data mining* dan *machine learning* sering dipakai untuk menyebutkan sesuatu yang sama. Bidang ini bersinggungan dengan ilmu probabilitas dan statistic kadang juga optimasi. *Machine learning* menjadi alat analisis dalam *data mining*. (Santoso, 2007)



Gambar 2.1 Posisi Data Mining (Santoso, 2007)

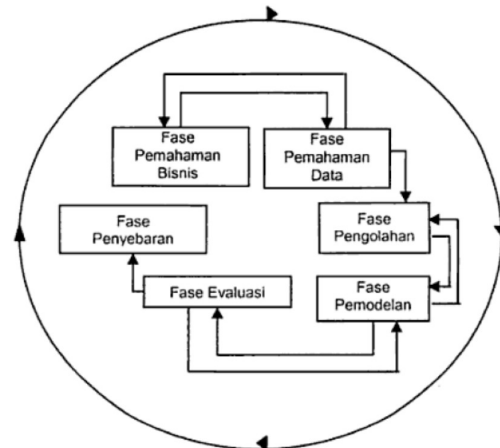
Para ahli berusaha menentukan posisi *data mining* diantara bidang-bidang yang lain. Hal ini dikarenakan ada kesamaan antara sebagian bahasan dalam *data mining* dengan bahasan di bidang lain. Memang tidak seratus persen sama, tetapi ada sejumlah kesamaan karakteristik dalam beberapa hal. Kesamaan bidang *data mining* dengan bidang statistik adalah penyampelan, estimasi, dan pengujian hipotesis. Kesamaan dengan kecerdasan buatan (*artificial intelligence*), pengenalan pola (*pattern recognition*), dan pembelajaran mesin (*machine learning*) adalah algoritma pencarian, teknik pemodelan, dan teori pembelajaran. Bidang lain yang mempengaruhi *data mining* adalah teknologi basis data, yang mendukung penyediaan penyimpanan yang efisien, pengindeksan, dan pemrosesan *query*. (Prasetyo, 2012)

2.3 Langkah-langkah dalam Data Mining

Cross-Industry Standard Process for Data Mining (CRISP-DM) yang dikembangkan tahun 1996 oleh analis dari beberapa industry seperti Daimler Chryslers, SPSS dan NCR. CRISP-DM menyediakan standar proses *data mining* sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian.

Dalam CRISP-DM, sebuah proyek *data mining* memiliki siklus hidup yang terbagi dalam enam fase

(Gambar II-2). Keseluruhan fase berurutan tersebut bersifat adaptif. Fase berikutnya dalam urutan bergantung kepada keluaran dari fase sebelumnya. Hubungan penting antarfase digambarkan dengan panah. Sebagai contoh, jika proses berada pada fase *modelling*. Berdasarkan pada perilaku dan karakteristik model, proses mungkin harus kembali kepada fase *data preparation* untuk perbaikan lebih lanjut terhadap data atau berpindah maju kepada fase *evaluation*.



Gambar 2.2. Fase CRISP-DM (Kusrini & Luthfi, 2009, p. 20)

Enam fase CRISP-DM (Larose, 2005)

1. Fase Pemahaman Bisnis (*Business Understanding Phase*)
 - a. Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
 - b. Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan *data mining*.
 - c. Menyiapkan strategi awal untuk mencapai tujuan.
2. Fase Pemahaman Data (*Data Understanding Phase*)
 - a. Mengumpulkan data.
 - b. Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.
 - c. Mengevaluasi kualitas data.
 - d. Jika diinginkan, pilih sebagian kecil group data yang mungkin mengandung pola dari permasalahan.
3. Fase Pengolahan Data (*Data Preparation Phase*)
 - a. Siapkan dari data awal, kumpulan data yang akan digunakan untuk keseluruhan fase berikutnya. Fase ini merupakan pekerjaan berat yang perlu dilaksanakan secara intensif.
 - b. Pilih kasus dan variabel yang ingin dianalisis dan yang sesuai analisis yang akan dilakukan.
 - c. Lakukan perubahan pada beberapa variabel jika dibutuhkan.
 - d. Siapkan data awal sehingga siap untuk perangkat pemodelan.

4. Fase Pemodelan (*Modeling Phase*)
 - a. Pilih dan aplikasikan teknik pemodelan yang sesuai.
 - b. Kalibrasi aturan model untuk mengoptimalkan hasil.
 - c. Perlu diperhatikan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan *data mining* yang sama.
 - d. Jika diperlukan, proses dapat kembali ke fase pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik *data mining* tertentu.
5. Fase Evaluasi (*Evaluation Phase*)
 - a. Mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarkan untuk pengguna.
 - b. Menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal.
 - c. Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
 - d. Mengambil keputusan berkaitan dengan penggunaan hasil dari *data mining*.
6. Fase Penyebaran (*Deployment Phase*)
 - a. Menggunakan model yang dihasilkan. Terbentuknya model tidak menandakan telah terselesaikannya proyek.
 - b. Contoh sederhana penyebaran : pembuatan laporan.
 - c. Contoh kompleks penyebaran : penerapan proses *data mining* secara parallel pada departemen lain.

2.4 Evaluasi dan Validasi Metode Klasifikasi Data Mining

Evaluasi adalah kunci ketika membuat aplikasi berbasis *data mining*. Ada berbagai macam dalam melakukan evaluasi. Jika kita memiliki data yang kita gunakan dalam proses pelatihan, maka tidak serta merta kita menjadikan data tersebut sebagai indikator keberhasilan aplikasi yang kita buat. Oleh karena itu, kita membutuhkan metode tertentu guna memprediksi performa berdasarkan eksperimen untuk berbagai macam data selain *data training* tersebut.

Confusion Matrix

Confusion matrix atau konfusi matrik merupakan tabel pencatat hasil kerja klasifikasi. Tabel II-1 merupakan contoh matrik konfusi yang melakukan klasifikasi masalah biner (dua kelas), hanya ada dua kelas, yaitu kelas 0 dan 1. Setiap sel f_{ij} dalam matriks menyatakan jumlah record/data dari kelas i yang prediksinya masuk ke kelas j . misalnya sel f_{11} adalah jumlah data dalam kelas 1 yang secara benar dipetakan ke kelas 1, dan f_{10} adalah data dalam kelas 1 yang dipetakan secara salah ke kelas 0.

Tabel 2.1. Matriks konfusi untuk klasifikasi dua kelas

Fij		Kelas hasil prediksi (j)	
		Kelas = 1	Kelas = 0
Kelas asli (i)	Kelas = 1	F11	F10
	Kelas = 0	F01	F00

Berdasarkan isi matriks konfusi, kita dapat mengetahui data dari masing-masing kelas yang diprediksi secara benar, yaitu $(f_{11} + f_{00})$, dan data yang diklasifikasikan secara salah, yaitu $(f_{10} + f_{01})$. Kuantitas matriks konfusi dapat diringkas menjadi dua nilai, yaitu akurasi dan laju error. Dengan mengetahui jumlah data yang diklasifikasikan secara benar, kita dapat mengetahui akurasi hasil prediksi, dan dengan mengetahui jumlah data yang diklasifikasikan secara salah, kita dapat mengetahui laju error dari prediksi yang dilakukan. Dua kuantitas ini digunakan sebagai metric kinerja klasifikasi.

Untuk menghitung akurasi digunakan formula :

$$\begin{aligned}
 \text{Akurasi} &= \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} \\
 &= \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}}
 \end{aligned}$$

Untuk menghitung laju error (kesalahan prediksi) digunakan formula :

$$\begin{aligned}
 \text{laju error} &= \frac{\text{Jumlah data yang diprediksi secara salah}}{\text{jumlah prediksi yang dilakukan}} \\
 &= \frac{f_{10} + f_{01}}{f_{11} + f_{10} + f_{01} + f_{00}}
 \end{aligned}$$

Semua algoritma klasifikasi berusaha membentuk model yang mempunyai akurasi tinggi (laju error rendah). Umumnya, model yang dibangun dapat memprediksi dengan benar pada semua data yang menjadi data latihnya, tetapi ketika model berhadapan dengan data uji, barulah kinerja model dari sebuah algoritma klasifikasi ditentukan.

2.5 Kurva Receiver Operating Characteristic (ROC)

Salah satu metode yang dapat dipakai untuk menghitung nilai kesalahan dan nilai kesuksesan suatu sistem adalah *Receiver Operating Characteristic* (ROC) dan nilai kinerjanya dapat diukur melalui perhitungan daerah dibawah kurva ROC yang disebut dengan *Area Under the Receiver Operating Characteristics Curve* (AUC).

Istilah ROC muncul pertama kali dalam deteksi suatu sinyal, apakah termasuk dalam *hit rate* ataukah alarm yang salah (*false alarm*). Terkadang *Data Mining* yang kita hasilkan akan diuji terhadap kesalahan positif atau negatif.

Kurva ROC adalah gambaran dua dimensi dari kinerja suatu pengklasifikasi. Suatu metode yang umum digunakan untuk menghitung nilai kinerja dari pengklasifikasi adalah dengan menghitung luas daerah dibawah kurva ROC, disebut dengan AUC. Karena AUC adalah luas kurva dari suatu persegi empat, maka nilainya selalu berada diantara 0 dan 1. Untuk suatu kurva ROC yang memadai, maka letaknya selalu berada di daerah sebelah atas dari garis diagonal (0,0) dan (1,1), sehingga tidak ada nilai AUC yang lebih kecil dari 0,5.

Tabel 2.2 Kategori Pengklasifikasi berdasarkan nilai AUC

Nilai AUC	Diklasifikasikan sebagai
0.90 – 1.00	Excellent
0.80 – 0.90	Good
0.70 - 0.80	Fair
0.60 – 0.70	Poor
0.50 – 0.60	Fail

Metode Support Vector Machine

Metode ini berakar dari teori pembelajaran statistik yang hasilnya sangat menjanjikan untuk memberikan hasil yang lebih baik daripada metode yang lain. Ide dasar SVM adalah memaksimalkan batas *hyperplane* (*maximal margin hyperplane*). (Prasetyo, 2012)

James Sanger Ronen Feldman menyebutkan bahwa *Support Vector Machine* pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep-konsep unggulan dalam bidang *pattern recognition*. SVM adalah metode *learning machine* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada input space.(Susilowati et al. n.d.)

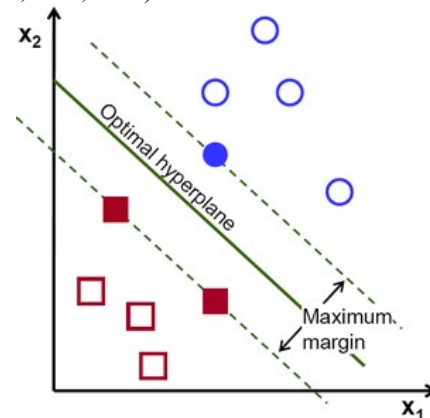
Hyperplane (batas keputusan) pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur margin *hyperplane* tersebut dan mencari titik maksimalnya. Margin adalah jarak antara *hyperplane* tersebut dengan data terdekat dari masing-masing kelas. Data yang paling dekat ini disebut *support vector*. Garis solid pada gambar x sebelah kanan menunjukkan *hyperplane* terbaik, yaitu yang terletak tepat pada tengah-tengah kedua kelas, sedangkan data lingkaran dan bujur sangkar yang dilewati garis batas margin (garis putus-putus) adalah *support vector*. Usaha untuk mencari lokasi *hyperplane* ini merupakan inti dari proses pelatihan pada SVM. (Prasetyo, 2012)

Meskipun waktu pelatihan SVM kebanyakan lambat, tetapi metode ini sangat akurat karena kemampuannya untuk menangani model-model nonlinier yang kompleks. SVM dapat digunakan untuk prediksi dan klasifikasi. (Widodo, et al., 2013)

2.6 Support Vector Machine (SVM) Linier

SVM memecahkan masalah klasifikasi dengan mencari *hyperplane* marjinal maksimum di mana ada

jumlah tak terbatas *hyperplanes* yang harus dicari mana yang terbaik. Secara intuitif, *hyperplane* dengan *margin* yang lebih besar lebih akurat dalam mengklasifikasikan data dibanding dengan *margin* yang lebih kecil. Inilah sebabnya mengapa (selama pembelajaran), mencari *hyperplane* dengan *margin* terbesar, dikenal dengan istilah *Maximum Marginal Hyperplane* (MMH). Untuk definisi margin, kita dapat mengatakan bahwa jarak terpendek dari *hyperplane* ke satu sisi margin adalah sama dengan jarak terpendek dari *hyperplane* yang ke sisi lain dari *margin*, di mana sis dari margin sejajar dengan *hyperplane* tersebut. (Widodo, et al., 2013)



Gambar 2.4 Margin hyperplane (Prasetyo, 2012)

Setiap data latih dinyatakan oleh (x_i, y_i) di mana $i=1,2,\dots,N$ dan $x_i = \{x_{i1}, x_{i2}, \dots, x_{iq}\}^T$ merupakan atribut (fitur) set untuk data latih ke- i , $y_i \in \{-1, +1\}$ menyatakan label kelas. *Hyperplane* klasifikasi linier SVM, seperti pada gambar diatas, dinotasikan dengan :

$$w \cdot x_i + b = 0 \dots \dots \dots (2.1)$$

w dan b adalah parameter model. $w \cdot x_i$ merupakan *inner product* dalam antara w dan x_i yang masuk ke dalam kelas -1 adalah data yang memenuhi pertidaksamaan berikut :

$$w \cdot x_i + b \leq -1 \dots \dots \dots (2.2)$$

Sementara data x_i yang masuk ke dalam kelas +1 adalah data yang memenuhi pertidaksamaan berikut :

$$w \cdot x_i + b \geq +1 \dots \dots \dots (2.3)$$

Sesuai dengan gambar II-4, jika ada data dalam kelas -1 (misalnya x_a) yang bertempat di *hyperplane*, persamaan (1) akan terpenuhi. Untuk data kelas -1 dinotasikan dengan

$$w \cdot x_a + b = 0 \dots \dots \dots (2.4)$$

Sementara kelas +1 (misal x_b) akan memenuhi persamaan

$$w \cdot x_b + b = 0 \dots \dots \dots (2.5)$$

Dengan menguragi persamaan (5) dan persamaan (6) didapatkan

$$w \cdot (x_a - x_b) = 0 \dots \dots \dots (2.6)$$

$x_b - x_a$ adalah *vector paralel* di posisi *hyperplane* dan diarahkan dari x_a ke x_b . Karena *inner product* dalam bernilai nol, arah w harus tegak lurus terhadap *hyperplane* (Gambar 2.4)

Dengan memberikan label -1 untuk kelas pertama dan +1 untuk kelas kedua, prediksi semua data uji akan menggunakan formula

$$y = \begin{cases} +1, & \text{jika } w \cdot z + b > 0 \\ -1, & \text{jika } w \cdot z + b < 0 \end{cases} \dots \dots \dots (2.7)$$

Sesuai dengan gambar, *hyperplane* untuk kelas -1 (garis putus-putus) adalah data pada *support vector* yang memenuhi persamaan

$$w \cdot x_a + b = -1 \dots \dots \dots (2.8)$$

sementara *hyperplane* kelas +1 (garis putus-putus) memenuhi persamaan

$$w \cdot x_b + b = +1 \dots \dots \dots (2.9)$$

dengan demikian, margin dapat dihitung dengan mengurangi persamaan (2.9) dan persamaan (2.8), didapatkan

$$w \cdot (x_b - x_a) = 2 \dots \dots \dots (2.10)$$

Margin *Hyperplane* diberikan oleh jarak antara dua *hyperplane* dari kelas tersebut. Notasi diatas diringkas menjadi

$$\|w\| \times d = 2 \text{ atau } d = \frac{2}{\|w\|} \dots \dots \dots (2.11)$$

3. Rancangan Sistem dan Aplikasi

Metode Penelitian

Penelitian diartikan sebagai proses pengumpulan dan analisis data yang dilakukan secara sistematis dan logis untuk mencapai tujuan-tujuan tertentu. Pengumpulan dan analisis data itu menggunakan metode-metode ilmiah, baik yang bersifat kuantitatif maupun kualitatif, eksperimental maupun non eksperimental, interaktif maupun non interaktif. (Sudaryono, 2015)

Penelitian berdasarkan fungsinya secara umum dan mendasar dapat dibedakan tiga macam penelitian, yaitu penelitian dasar, penelitian terapan dan penelitian evaluasi. (Sudaryono, 2015)

Penelitian prediksi kelulusan mahasiswa berdasarkan fungsinya termasuk kedalam kelompok Penelitian Terapan (*applied research*). Hasil dari penelitian dapat langsung diterapkan untuk memecahkan permasalahan yang dihadapi.

Fokus dari penelitian ini menerapkan model *Support Vector Machine* (SVM) untuk memprediksi kelulusan mahasiswa di STMIK Cikarang. Diharapkan hasil dari penelitian ini bisa membantu STMIK Cikarang mengurangi jumlah mahasiswa yang tidak lulus.

Sampling/Metode Pemilihan Sampel

Sampel yang baik merupakan sampel yang akurat dan tepat. Sampel yang tidak akurat dan tidak tepat akan memberikan kesimpulan riset yang tidak diharapkan atau dapat menghasilkan kesimpulan salah dan menyesatkan. (Jogiyanto, 2008)

Adapun pada penelitian ini metode pemilihan sampel untuk jenis data primer, peneliti menggunakan metode *simple random sampling*. Data diambil dari data mahasiswa angkatan 2006 sampai dengan 2010.

Metode Pengumpulan Data

Metode pengumpulan data adalah cara atau teknik yang dapat digunakan oleh peneliti untuk mengumpulkan data. Pengumpulan data dalam penelitian dimaksudkan untuk memperoleh bahan, keterangan, kenyataan, dan informasi yang dapat dipercaya. (Sudaryono, 2015)

Pada penelitian ini, data diperoleh dengan metode wawancara, observasi lapangan dan studi literatur. Wawancara dan observasi lapangan untuk mendapatkan data primer, sedangkan studi literatur untuk mendapatkan data sekunder.

Wawancara akan dilakukan dengan Ketua STMIK Cikarang, Wakil Ketua Bidang Akademik, dan Wakil Ketua Bidang Kemahasiswaan. Hal-hal yang akan ditanyakan berhubungan dengan proses kelulusan dan kebijakan berhubungan dengan kelulusan mahasiswa.

Observasi lapangan dilakukan dengan mengumpulkan dokumen-dokumen terkait dengan proses kelulusan di STMIK Cikarang dan dokumen rancangan Sistem Informasi Akademik. Observasi lapangan akan banyak dilakukan di Pusdat STMIK Cikarang.

Studi pustaka dilakukan dengan mempelajari, meneliti dan membaca buku, jurnal, skripsi atau tesis yang berhubungan dengan *data mining* dan kegiatan akademik.

Instrumentasi Penelitian

Metode pengumpulan data yang peneliti lakukan seperti yang telah disebutkan diatas antara lain melalui observasi dan wawancara, maka instrumennya adalah peneliti sendiri.

Teknik Analisis Data

Proses analisis dilakukan terhadap hasil tahapan pengumpulan data dengan wawancara, observasi dan studi pustaka untuk mendapatkan spesifikasi kebutuhan sistem yang akan dikembangkan.

Data yang didapatkan dari hasil wawancara dengan pimpinan STMIK Cikarang, diharapkan dapat digunakan untuk merumuskan permasalahan data mining pada fase Pemahaman Bisnis (*Business Understanding Phase*)

Data yang didapatkan dari hasil observasi di STMIK Cikarang, diharapkan bisa digunakan dalam fase pemahaman data. Dengan memilih data yang didapatkan dari sistem informasi akademik, akan dipilih data dan dilakukan proses *cleansing* data untuk dapat dilakukan proses *data mining*.

Langkah-langkah Penelitian

Berdasarkan pada *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang dikembangkan oleh analis dari beberapa industri, maka penelitian ini mengadopsi model tersebut dengan 6 tahapan penelitian sebagai berikut:

1. Fase Pemahaman Bisnis (*Business Understanding Phase*)
2. Fase Pemahaman Data (*Data Understanding Phase*)
3. Fase Pengolahan Data (*Data Preparation Phase*)
4. Fase Pemodelan (*Modeling Phase*)
5. Fase Evaluasi (*Evaluation Phase*)
6. Fase Penyebaran (*Deployment Phase*)

Fase Pemahaman Bisnis (*Business Understanding Phase*)

Permasalahan yang ada di STMIK Cikarang saat ini adalah tingkat kelulusan mahasiswa yang masih tinggi setiap tahunnya. Berangkat dari masalah itu, pada fase pemahaman bisnis ini, peneliti mengumpulkan data untuk mendapatkan informasi mengenai pentingnya mempelajari data yang telah lampau menggunakan teknik *data mining*. Langkah ini dilakukan dengan cara melakukan wawancara dan observasi di kampus STMIK Cikarang.

Wawancara akan dilakukan dengan Ketua STMIK Cikarang dan Puket I bidang Akademik. Dua pejabat ini peneliti nilai sebagai pihak yang berkepentingan secara langsung terhadap tingkat kelulusan mahasiswa.

Fase Pemahaman Data (*Data Understanding Phase*)

Tahap pemahaman data, akan dilakukan pemilahan tabel-tabel yang terkait dengan kelulusan mahasiswa. Tabel-tabel tersebut diambil dari *database system* informasi akademik yang sudah berjalan selama ini. Dari tabel terpilih, diambil data pada range waktu dari angkatan 2006 sampai dengan angkatan 2011.

Fase pemahaman data ini peneliti akan melakukan observasi di bagian puskom STMIK Cikarang. Dengan metode wawancara ke kepala puskom dan observasi ke Sistem Akademik.

Fase Pengolahan Data (*Data Preparation Phase*)

Hasil dari fase pemahaman data, akan menjadi modal peneliti untuk masuk ke fase pengolahan data. Pada fase ini peneliti masih bekerja sama dengan bagian puskom untuk melakukan pengolahan data.

Tahap ini meliputi semua kegiatan untuk membangun *dataset* akhir (data yang akan diproses pada tahap pemodelan/*modeling*) dari data mentah. Tahap ini dapat diulang beberapa kali. Pada tahap ini juga mencakup pemilihan tabel, *record*, dan atribut-atribut data, termasuk proses pembersihan dan transformasi data untuk kemudian dijadikan masukan dalam tahap pemodelan (*modeling*).

Fase Pemodelan (*Modeling Phase*)

Dalam tahap ini akan dilakukan penerapan algoritma *Support Vector Machine* (SVM). Data yang diperoleh dari fase pengolahan data akan digunakan dalam proses ini. Akan dibuat dua dataset untuk dibuat model *data mining* dan dibandingkan hasil evaluasinya.

Fase Evaluasi (*Evaluation Phase*)

Pada tahap ini akan dilakukan evaluasi terhadap keefektifan dan kualitas model data mining menggunakan algoritma *Support Vector Machine* (SVM) sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang ditetapkan pada fase awal (*Business Understanding*).

Metode yang digunakan dalam fase ini adalah *Confusion Matrix* dan Kurva ROC. Dari data *Confusion matrix* dapat dihitung akurasi dan laju error data model *data mining*. Dan dari Kurva ROC dapat diketahui nilai AUC. Dari nilai-nilai tersebut, dapat diketahui dan dipilih model *data mining* yang akan dibuat menjadi aplikasi *data mining*.

Fase Penyebaran (*Deployment Phase*)

Tahap *deployment* dilakukan dengan pembuatan aplikasi data mining dengan menggunakan model yang terbentuk pada proses sebelumnya sehingga aplikasi bisa digunakan untuk *data training* dari data yang berbeda.

Pemodelan aplikasi menggunakan UML dengan *Use Case Diagram* dan *Activity Diagram*. Sedangkan aplikasi akan dibangun dengan menggunakan software *Microsoft SQL Server 2008R2* untuk menangani databasenya dan *Microsoft Visual Studio 2010* untuk membuat aplikasi *front end*. Untuk pengujian aplikasi, akan digunakan metode pengujian kotak hitam (*black box testing*).

Hasil dari fase ini diharapkan sebuah aplikasi *data mining* yang siap pakai dan teruji dan juga bermanfaat untuk STMIK Cikarang.

4. PEMBAHASAN DAN HASIL PENELITIAN

Fase Pemahaman Bisnis (*Business Understanding Phase*)

Dari hasil wawancara dengan Ketua Penjaminan Mutu STMIK Cikarang dan observasi yang dilakukan, didapati masalah-masalah dan harapan sebagai berikut :

1. Sistem yang ada saat ini belum dapat mengetahui mahasiswa berpotensi tidak lulus, sehingga tindakan pencegahan tidak dapat dilakukan.
2. Belum ada pemanfaatan data lulusan yang telah lampau untuk digunakan mempelajari trend kelulusan.
3. Perlunya dibuat sebuah sistem yang dapat mengetahui mahasiswa berpotensi tidak lulus yang bisa menjadi bahan masukan untuk pimpinan dalam menentukan kebijakan terkait dengan tindakan pencegahan.

Fase Pemahaman Data (Data Understanding Phase)

Observasi dilakukan pada Sistem Informasi Akademik STMIK Cikarang dan Sistem Feeder PDDikti STMIK Cikarang. Hal ini dilakukan untuk dapat mengetahui dan menggali data lampau yang akan digunakan untuk proses *datamining*.

Dari hasil observasi tersebut, didapatkan tabel-tabel yang berhubungan dengan data kelulusan adalah:

1. Mahasiswa
2. Nilai Perkuliahan
3. Kelulusan
4. Biaya Kuliah

Data-data tersebut dengan menggunakan tools Microsoft Visual Foxpro, Microsoft SQL Server dan Microsoft Excel untuk dapat menghasilkan dataset.

Fase Pengolahan Data (Data Preparation Phase)

Untuk mendapatkan dataset yang diinginkan, maka proses berikutnya akan dilakukan pemilihan atribut yang akan digunakan dalam proses pemodelan. Perintah-perintah query seleksi dan join dilakukan disini.

Data yang digunakan dalam penelitian ini adalah data sekunder. Data diperoleh dari objek penelitian, yaitu STMIK Cikarang. Dalam penelitian ini akan dibuat 2 Dataset dengan jumlah parameter yang berbeda.

Untuk Dataset yang pertama, parameter *input* sejumlah lima buah. Sedangkan parameter *output*-nya adalah satu buah. Hal itu dapat dilihat pada tabel di berikut ini.

Tabel 4.1. Parameter Dataset A

No	Nama Parameter	Keterangan Nama Parameter	Uraian Paramater
1.	IPK	Indeks Prestasi mahasiswa	IPK Mahasiswa bersangkutan di Semester 4
2.	Usia	Usia mahasiswa	Usia mahasiswa ketika mendaftar
3.	JK	Jenis Kelamin Mahasiswa	1 : Perempuan 2 : Laki-laki
5.	KodeProdi	Kode Program Studi	1: Teknik Informatika 2: Sistem Informasi 3 : Manajemen Informatika 4:Komputerisasi Akuntansi
6.	Status_Lulus		Lulus dan Tidak Lulus

Dataset yang kedua, hampir sama dengan dataset yang kedua, hanya saja ditambahkan parameter Sisa Angsuran.

Tabel 4.1. Parameter Dataset B

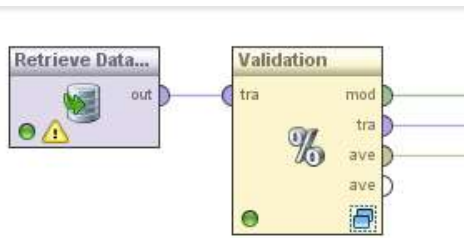
N	Nama Parameter	Ket Nama Parameter	Uraian Paramater
1.	IPK	Indeks Prestasi mahasiswa	IPK Mahasiswa bersangkutan di Semester 4
2.	Usia	Usia mahasiswa	Usia mahasiswa ketika mendaftar
3.	JK	Jenis Kelamin Mahasiswa	1 : Perempuan 2 : Laki-laki
5.	KodeProdi	Kode Program Studi	1: Teknik Informatika 2: Sistem Informasi 3 : Manajemen Informatika 4:Komputerisasi Akuntansi
	sisaAngsuran	Sisa tunggakan angsuran pada akhir semester 4	
6.	Status_Lulus		Lulus dan Tidak Lulus

Proses pengolahan data menghasilkan 758 records untuk dataset A dan dataset B yang selanjutnya akan dinormalisasi dan diolah dengan *aplikasi Rapid Miner 5.2*. Dengan menggunakan Metode Pemilihan Sampel *Stratified Sampling*, Dataset A dan Dataset B tersebut digunakan sebagai *data training* dan *data testing*. Dengan menggunakan *10-fold cross validation*, data akan diolah untuk menghasilkan model *data mining*.

Fase Pemodelan Data (Modelling Phase)

Model *data mining* dalam penelitian ini dibuat dengan menggunakan perangkat lunak adalah *Rapid Miner 5.2*. Algoritma *Support Vector Machine (SVM)* dijalankan dengan menggunakan perangkat lunak tersebut.

Model dibuat dengan sumber data dari Dataset A dan dataset B. Dataset A dan Dataset B terlebih dahulu diimport ke dalam aplikasi *Rapid Miner*. Dengan cara yang sama, berikut adalah model data mining menggunakan *Rapid Miner 5.3*.



Gambar 4.1 Model Data Mining di Rapidminer

Langkah selanjutnya adalah penentuan model algoritmanya, dalam hal ini adalah *Support Vector Machine* (SVM). Pada tahap ini sekaligus ditambahkan tools untuk mengetahui performasi dari model. Proses ini dilakukan untuk masing-masing dataset yang telah dibuat.

Dari proses pemodelan ini, didapatkan hasil nilai weight dari masing-masing parameter yang diberikan. Nilai weight inilah yang digunakan dalam proses prediksi.

Nilai w (*Weight*) dari masing-masing parameter pada Dataset 1 dan Dataset 2 sebagai berikut :

Tabel Nilai w untuk Dataset 1

Tabel 4.1 Nilai w (*weight*) untuk dataset A

No	Parameter	Nilai w (<i>Weight</i>)
1.	Kode Prodi	0.015
2.	Kode Jenis Kelamin	-0.167
3.	Umur	0.006
4.	IPK	0.384
5.	Bias	-1.182

Tabel 4.2. Nilai w (*weight*) untuk dataset B

No	Parameter	Nilai w (<i>Weight</i>)
1.	Kode Prodi	0.021
2.	Kode Jenis Kelamin	-0.185
3.	Umur	0.013
4.	IPK	0.371
5.	Sisa Angsuran	0.236
6.	Bias	0.852

Fase Evaluasi (*Evaluation Phase*)

Evaluasi dari model *data mining* diatas dilakukan dengan menghitung Akurasi dan laju error dari percobaan terhadap 2 dataset diatas. Metode evaluasi yang dilakukan adalah *Confusion matrik* dan Kurva *Receiver Operating Characteristic*.

Confusion Matrix

Data *Confusion Matrix* untuk model Dataset A yang didapatkan dari Aplikasi *RapidMiner* dapat dilihat pada tabel berikut :

Tabel 4.3 Data *Confusin Matrix* Dataset A

Komponen	True Lulus	True Tidak Lulus
Pred. Lulus	445	105
Pred. Tidak Lulus	29	169

Dari hasil diatas, dapat dihitung nilai akurasi dan laju error dari Dataset A sebagai berikut :

$$\text{Akurasi} = \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} \times 100\%$$

$$= \frac{445 + 169}{445 + 29 + 105 + 169} \times 100\% = \frac{614}{748} \times 100\% = 82.32\%$$

$$\text{laju error} = \frac{\text{Jumlah data yang diprediksi secara salah}}{\text{jumlah prediksi yang dilakukan}} \times 100\%$$

$$= \frac{29 + 105}{455 + 29 + 105 + 169} \times 100\% = \frac{134}{758} \times 100\% = 17.67\%$$

Data *Confusion Matrix* untuk model Dataset B yang didapatkan dari Aplikasi *RapidMiner* dapat dilihat pada tabel berikut :

Tabel 4.4. Data *Confusin Matrix* Dataset B

Komponen	True Lulus	True Tidak Lulus
Pred. Lulus	453	102
Pred. Tidak Lulus	31	172

Dari hasil diatas, dapat dihitung nilai akurasi dan laju error dari Dataset A sebagai berikut :

$$\text{Akurasi} = \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} \times 100\%$$

$$= \frac{453 + 172}{453 + 31 + 102 + 172} \times 100\% = \frac{625}{758} \times 100\% = 82.45\%$$

$$\text{laju error} = \frac{\text{Jumlah data yang diprediksi secara salah}}{\text{jumlah prediksi yang dilakukan}} \times 100\%$$

$$\begin{aligned}
 &= \frac{31 + 102}{453 + 31 + 102 + 172} \times 100\% \\
 &= \frac{133}{758} \times 100\% = 17.54\%
 \end{aligned}$$

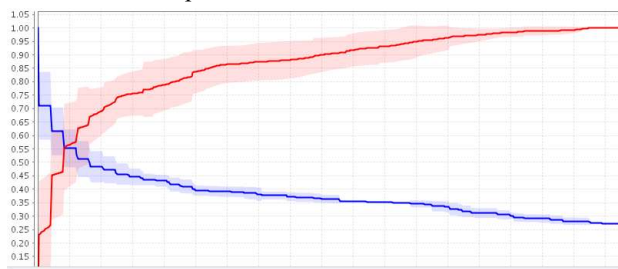
Secara keseluruhan, tabel berikut membandingkan hasil percobaan pembuatan model data mining dengan 2 dataset yang berbeda. Dapat dilihat bahwa dataset B mempunyai nilai Akurasi lebih tinggi dibandingkan dengan dataset a. Demikian juga dengan nilai laju error, Dataset B memiliki laju error lebih rendah dibandingkan dengan Dataset A. sehingga dapat disimpulkan bahwa model data mining dari Dataset B lebih baik dibandingkan dengan model data mining dataset A.

Tabel 4.5. Nilai Akurasi dan Laju Error

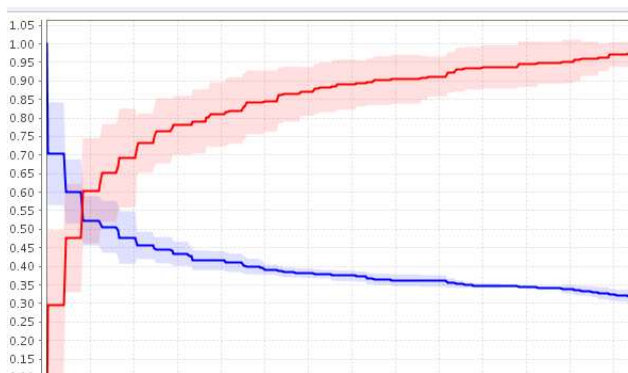
Komponen	Dataset A	Dataset B
Akurasi	82.32%	82.45%
Laju Error	17.67%	17.54%

Kurva Receiver Operation Characteristic (ROC)

Selain *confusion matrix*, kurva *Receiver Operating Characteristic (ROC)* dihasilkan oleh *Rapid Miner*. Kurva tersebut dapat dilihat di bawah ini.



Gambar 4.3. Kurva ROC untuk dataset A



Gambar 4.4 Kurva ROC untuk dataset B

Dengan menggunakan *tools rapidminer*, dapat diketahui nilai AUC untuk masing-masing dataset.

Tabel 4.6. Nilai AUC dari hasil percobaan dengan *rapidminer*

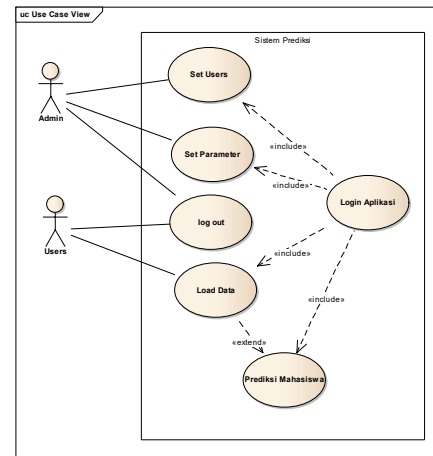
Data	Nilai AUC
Dataset A	0.881
Dataset B	0.886

Tabel 4.6 membandingkan nilai AUC hasil percobaan pembuatan model data mining dengan 2 dataset yang berbeda. Dapat dilihat bahwa dataset B mempunyai nilai AUC lebih tinggi dibandingkan dengan dataset A. Sehingga dapat disimpulkan bahwa model data mining dari Dataset B lebih baik dibandingkan dengan model data mining dataset A walaupun keduanya masuk kedalam kategori **GOOD**.

Fase Penyebaran (Deployment Phase)

Dari hasil percobaan pembuat model *data mining* dengan *rapidminer* didapati bahwa model Dataset B lebih baik, langkah selanjutnya adalah fase penyebaran (*deployment Phase*). Pada fase ini, tujuan akhirnya adalah membangun prototype aplikasi *data mining*. Langkah dalam pembuatan prototype ini adalah membuat analisa sistem dengan menggunakan UML, mendesain database dan selanjutnya membuat aplikasinya.

Analisa Sistem Use Case Diagram



Gambar 4.5 Use Case Diagram

Use case diagram dari aplikasi prediksi kelulusan mahasiswa ini terdiri dari 2 actor. Actor yang pertama adalah admin yang bertugas untuk mengatur dan mendaftarkan user yang dapat menggunakan aplikasi ini, dan juga untuk melakukan maintenance data nilai parameter parameter yang digunakan dalam proses prediksi menggunakan algoritma *Support Vector Machine (SVM)*.

Actor yang kedua adalah users dari aplikasi ini. Users disini bisa Ketua STMIK Cikarang, Kaprodi untuk Program Studi yang ada di STMIK Cikarang, juga bagian lain yang

sekiranya berkepentingan untuk mengakses aplikasi ini. Yang bisa dilakukan oleh *actor* ini adalah memuat data mahasiswa, dan melakukan prediksi terhadap mahasiswa yang terpilih tersebut.

Desain Tabel

1. Tabel User

Column Name	Data Type	Allow Nulls
nik	varchar(20)	<input type="checkbox"/>
nama	varchar(35)	<input checked="" type="checkbox"/>
Password	varchar(15)	<input checked="" type="checkbox"/>
typeUser	int	<input checked="" type="checkbox"/>
expiredDate	smalldatetime	<input checked="" type="checkbox"/>

Gambar 4.6 Desain tabel tbUser

Tabel user digunakan untuk memanipulasi data user yang bisa mengakses data aplikasi. NIK adalah nomor induk karyawan yang digunakan sebagai primary key dari tabel. Nama dan password digunakan untuk menyimpan data bersangkutan. Type user bertipe integer akan bernilai 1 dan 2. Angka 1 digunakan untuk mewakili type user admin, dan angka 2 digunakan untuk mewakili type user operator.

2. Tabel Parameter

Column Name	Data Type	Allow Nulls
id	int	<input type="checkbox"/>
ParameterName	varchar(50)	<input checked="" type="checkbox"/>
ParameterWeight	decimal(18, 7)	<input checked="" type="checkbox"/>

Gambar 4.7 Desain Tabel Parameter

Tabel parameter digunakan untuk menyimpan data-data parameter aplikasi *data mining*. Data ini akan menyimpan nama parameter yang didapat dari model yang telah dibuat. Id menggunakan type data integer dan dibuat auto increment. Parameter Name diambil dari nama parameter dari *rapidminer* dan Parameter Weight menyimpan data bobot yang juga didapatkan dari *rapidminer*.

3. Tabel Mahasiswa

Column Name	Data Type	Allow Nulls
nim	char(10)	<input type="checkbox"/>
nama	varchar(50)	<input checked="" type="checkbox"/>
jeniskel	char(1)	<input checked="" type="checkbox"/>
prodi	char(2)	<input checked="" type="checkbox"/>
tahunmasuk	char(4)	<input checked="" type="checkbox"/>
tanggallahir	date	<input checked="" type="checkbox"/>
ipk	decimal(18, 4)	<input checked="" type="checkbox"/>

Gambar 4.8 Desain Tabel Mahasiswa

Tabel mahasiswa akan menyimpan data-data mahasiswa, dimulai dari NIM (Nomor Induk Mahasiswa), Nama Mahasiswa, Jenis Kelamin, Program Studi, Tahun Masuk, Tanggal Lahir dan IPK.

Desain Aplikasi

1. Form Login

Gambar 4.9. Form Login

Form login merupakan form yang pertama kali muncul ketika aplikasi pertama kali dijalankan. Terdapat isian username dan password yang harus disikan oleh user.

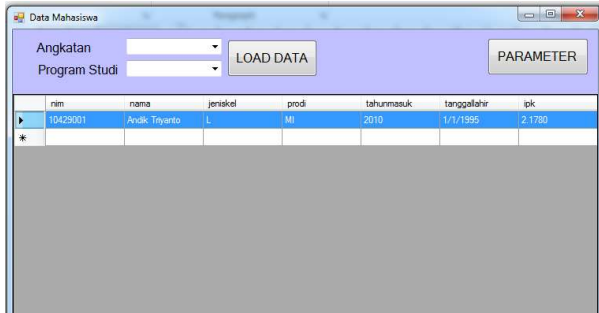
2. Form Parameter

parametername	parameterweight
KODEPRODI	0.0210000
USIAMHS	-0.1850000
KODEJK	0.0130000
IPK	0.3710000
ANGSURAN	0.2360000
BIAS	0.8520000

Gambar 4.10 Form Parameter

Form parameter menyimpan data-data nilai bobot dari masing-masing parameter. Nilai ini secara default sudah diisikan oleh sistem, user hanya bisa melakukan update data terhadap nilai weight jika aplikasi akan menggunakan model *data mining* yang berbeda.

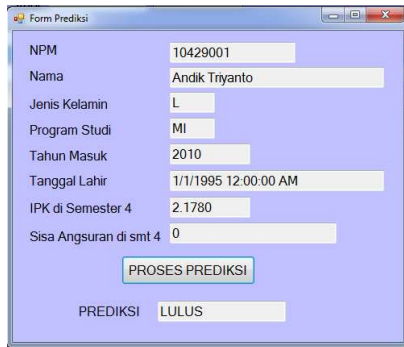
3. Form Mahasiswa



Gambar 4.11 Form Mahasiswa

Form mahasiswa digunakan untuk menampilkan data-data mahasiswa yang diambil dari sistem informasi akademik. Dari data mahasiswa ini selanjutnya akan dilakukan prediksi apakah mahasiswa tersebut lulus atau ada kemungkinan tidak lulus.

4. Form Prediksi



Gambar 4.12 Form Prediksi

Dalam form inilah sebenarnya proses prediksi dilakukan. Didalam prosesnya, ada perhitungan-perhitungan berdasarkan model yang sudah terbentuk sebelumnya. Hasil prediksi akan menunjukkan nilai lulus atau tidak lulus.

Implikasi Penelitian

Aspek Sistem

Untuk dapat mengoperasikan Sistem Prediksi Kelulusan Mahasiswa dengan baik, perlu dibuat suatu rancangan implementasi kebutuhan minimal sumber daya teknologi yang meliputi kebutuhan *software* dan *hardware*. Rancangan tersebut dapat dilihat pada tabel berikut:

Tabel 4.7 Rancangan Kebutuhan *Software* dan *hardware*

Nama	Kebutuhan Minimal
Sistem Operasi	Windows XP Profesional
Memory	512 MB
Graphic Card	64 MB
Harddisk	10 GB

Aspek Manajerial

Sistem Prediksi Kelulusan Mahasiswa diharapkan dapat meningkatkan prosentase kelulusan di STMIK Cikarang. Oleh karena itu, pelatihan yang cukup harus diberikan kepada sumber daya yang akan mengoperasikan sistem ini.

Aspek Penelitian Lanjut

Disadari bahwa penelitian ini masih banyak kekurangan. Oleh karena itu, perlu dilakukan penelitian lebih lanjut untuk melengkapi kekurangan yang ada terutama demi kesempurnaan model prediksi kelulusan mahasiswa yang telah dibuat sehingga dapat diperoleh tingkat akurasi prediksi yang optimal.

Agar mendapat tingkat akurasi yang optimal, akan lebih baik apabila algoritma *Support Vector Machine* dibandingkan dengan metode algoritma yang lain.

5. PENUTUP

Kesimpulan

Dari serangkaian pengolahan data dan analisa yang dilakukan pada penelitian / tesis ini dapat ditarik kesimpulan sebagai berikut :

- 1) Dengan menggunakan tugas *data mining* klasifikasi, algoritma *Support Vector Machine* (SVM) dapat diterapkan untuk prediksi kelulusan mahasiswa di STMIK Cikarang.
- 2) Dengan menggunakan tools *rapidminer* dapat dibuat model *datamining* yang menjadi dasar dari pembuatan aplikasi Prediksi Kelulusan yang dibuat dengan menggunakan tools Microsoft Visual C#.

Saran

Dalam rangka implementasi Sistem Prediksi Kelulusan Mahasiswa di STMIK Cikarang, ada beberapa saran saran yang dapat diberikan. Hal itu dapat dilihat pada uraian sebagai berikut.

- 1) Agar dapat berjalan secara optimal, Sistem Prediksi Kelulusan Mahasiswa diimplementasikan pada perangkat minimal sesuai dengan Rancangan Kebutuhan *Software* dan *Hardware* yang telah dibuat. Pelatihan yang cukup sebaiknya diberikan kepada sumber daya manusia yang akan mengoperasikan sistem tersebut. Selain itu, agar mendapat tingkat akurasi prediksi kelulusan mahasiswa yang optimal, pada penelitian yang akan datang, lebih baik apabila algoritma *Support Vector Machine* (SVM) yang sudah digunakan pada penelitian ini dibandingkan lagi dengan algoritma yang lain.
- 2) Sistem Prediksi kelulusan mahasiswa diharapkan diimplementasikan di STMIK Cikarang. Evaluasi dan perbaikan terus dilakukan, agar sistem tersebut dapat berjalan secara optimal.

DAFTAR PUSTAKA

- AS, R. & M. Shalahuddin, 2013. *Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek*. Bandung: Penerbit Informatika.
- Chapple, M., 2008. *Microsoft SQL Server 2008 for Dummies*,
- Handayani, L. & Fitriandini, 2010. Prediksi Kebangkrutan Perusahaan menggunakan Support Vector Machine (SVM).
- Hermawati, F. A., 2013. *Pengolahan Citra Digital Konsep dan Teori*. Yogyakarta: Penerbit Andi.
- Jogiyanto, 2008. *Metodologi Penelitian Sistem Informasi*. Yogyakarta: Penerbit Andi.
- Kusrini & Luthfi, E. T., 2009. *Algoritma Data Mining*. Yogyakarta: Penerbit Andi.
- Larose, daniel T., 2005. *Discovering Knowledge in Data (an Introduction to Data Mining)*,
- Nurhayati, S., Luthfi, E.T. & Papua, U.Y., 2011. Prediksi Mahasiswa Drop Out Menggunakan Metode Support Vector. *Prediksi menggunakan SVM*, 3(6), pp.82–93.
- Prasetyo, E., 2012. *Data Mining Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: Penerbit Andi.
- Prasetyo, E., 2014. *Data Mining Mengolah Data Menjadi Informasi menggunakan Matlab*. Yogyakarta: Penerbit Andi.
- Ramadhan, 2007. *Perancangan Data Warehouse dengan Microsoft SQL Server 2005*. s.l.:s.n.
- Rumbaugh, J., Jacobson, I. & Booch, G., 2004. *The Unified Modeling Language Reference Manual*,
- Sholih, 2006. *Pemodelan Sistem Informasi Beroientasi Object dengan UML*. Pertama ed. Yogyakarta: Graha Ilmu.
- Sudaryono, 2015. *Metodologi Riset di Bidang TI*. Yogyakarta: Penerbit Andi.
- Susilowati, E., Sabariah, M.K. & Gozali, A.A., Implementasi Metode Support Vector Machine untuk melakukan Klasifikasi Kemacetan Lalu Lintas pada Twitter. , pp.1–7.
- Widodo, P. P., 2011. *Menggunakan UML*. Bandung: Penerbit Informatika.
- Widodo, P. P., Handayanto, T. R. & Herlawati, 2013. *Penerapan Data Mining dengan Matlab*. Bandung: Rekayasa Sains.
- Ugm, F., 2014. Penerapan Metode Support Vector Machine pada Sistem Deteksi Intrusi secara Real-time. , 8(1), pp.13–24.